1. **Паспорт Образовательной программы**

**«Введение в Big Data: анализ естественного языка и взаимодействия пользователей»**

|  |  |
| --- | --- |
| **Версия программы** | 2 |
| **Дата Версии** | 15**.**10**.**2020 |

1. **Сведения о Провайдере**

|  |  |  |
| --- | --- | --- |
| 1.1 | Провайдер | Томский государственный университет |
| 1.2 | Логотип образовательной организации |  |
| 1.3 | Провайдер ИНН | 7018012970 |
| 1.4 | Ответственный за программу ФИО | Гойко Вячеслав Леонидович |
| 1.5 | Ответственный должность | Директор, Центр прикладного анализа больших данных |
| 1.6 | Ответственный Телефон | +7 952 888-91-24 |
| 1.7 | Ответственный Е-mail | goiko@data.tsu.ru |

1. **Основные Данные**

|  |  |  |
| --- | --- | --- |
| **№** | **Название** | **Описание** |
| 2.1 | Название программы | Введение в Big Data: анализ естественного языка и взаимодействия пользователей |
| 2.2 | Ссылка на страницу программы | <https://moodle.ido.tsu.ru/course/view.php?id=1197> |
| 2.3 | Формат обучения | Онлайн |
|  | Подтверждение от ОО наличия возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа | ОО подтверждает наличие возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа.  В Институте дистанционного образования ТГУ создана учебно-методическая и технологическая база системы дистанционного образования ТГУ, основанная на мультимедийных и web-технологиях. Институт обладает современным компьютерным оборудованием и лицензионным программным обеспечением, необходимым для подготовки и создания мультимедиакурсов, методической и технологической поддержки учебного процесса, подробнее https://ido.tsu.ru/about/ |
| 2.4 | Уровень сложности | Базовый |
| 2.5 | Количество академических часов | **72** часа |
|  | Практикоориентированный характер образовательной программы: не менее 50 % трудоёмкости учебной деятельности отведено практическим занятиям и (или) выполнению практических заданий в режиме самостоятельной работы (кол-во академических часов) | Практикоориентированный характер образовательной программы: 36 часов трудоёмкости учебной деятельности отведено практическим занятиям |
| 2.6 | Стоимость обучения одного обучающегося по образовательной программе, а также предоставление ссылок на 3 (три) аналогичные образовательные программы иных организаций, осуществляющих обучение, для оценки объективности стоимости или обоснование уникальности представленной образовательной программы в случае отсутствия аналогичных образовательных программ на рынке образовательных услуг | 25 000  [**https://www.bigdataschool.ru/**](https://www.bigdataschool.ru/)  [**https://www.bigdataschool.ru/about/price.html**](https://www.bigdataschool.ru/about/price.html)  [**https://www.infosystems.ru/courses/avtorskie\_unikalnye/mashinnoe\_obuchenie\_big\_data/**](https://www.infosystems.ru/courses/avtorskie_unikalnye/mashinnoe_obuchenie_big_data/) |
| 2.7 | Минимальное количество человек на курсе | 30 |
| 2.8 | Максимальное количество человек на курсе | 60 |
| 2.9 | Данные о количестве слушателей, ранее успешно прошедших обучение по образовательной программе | Нет |
| 2.10 | Формы аттестации | Защита проекта |
|  | Указание на область реализации компетенций цифровой экономики, к которой в большей степени относится образовательная программа, в соответствии с Перечнем областей | Большие данные |

1. **Аннотация программы**

Программа ориентирована на формирование комплексных знаний и компетенций, необходимых для профессиональной деятельности и проведения научно-исследовательской работы с применением технологий Big Data: NLP (обработки естественного языка – natural language processing) и анализа взаимодействия (social network analysis).

Программа рассчитана на представителей из разных областей: психологов, социологов, программистов, политологов, лингвистов, менеджеров и др. и направлен на формирование и совершенствование следующих цифровых компетенций исследователей:

* способность формировать стратегию и методологию социальных исследований с применением технологий Big Data;
* способность обоснованно выбирать и эффективно использовать технологии, методы и инструменты исследования;
* владение методами и инструментальными средствами социально-сетевого анализа;
* владение методами и программными инструментами текстовой аналитики для обработки текстов на естественном языке;
* способность анализировать и интерпретировать полученные результаты.

Объем модуля – 2 з.е., 72 часа (34 аудиторных (онлайн) и 38 СРС)

1. ШАБЛОН ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ (ДПО)

Титульный лист программы

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

Дополнительная профессиональная программа повышения квалификации

«Введение в Big Data: анализ естественного языка и взаимодействия пользователей»

72 час.

****

**ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

**1.Цель программы**

Развитие и углубление способностей и навыков в области анализа данных и проведения научно-исследовательской работы с применением технологий анализа больших данных.

**2.Планируемые результаты обучения:**

2.1.Знание (осведомленность в областях)

2.1.1. Алгоритмы социально-сетевого анализа

2.1.2. Алгоритмы обработки естественного языка

2.2. Умение (способность к деятельности)

2.2.1. Способность формировать стратегию и методологию социальных исследований с применением технологий Big Data;

2.2.2. Способность обоснованно выбирать и эффективно использовать технологии, методы и средства проведений исследования;

2.2.3. Способность собирать и обрабатывать большие массивы гетерогенных данных;

2.3.Навыки (использование конкретных инструментов)

2.3.1. Владение методами и инструментальными средствами сетевого анализа данных

2.3.2. Владение средствами интеллектуального анализа текстов

**3.Категория слушателей** (возможно заполнение не всех полей)

* 1. Образование:
* лица, имеющие среднее профессиональное и (или) высшее образование, наличие указанного образования должно подтверждаться документом государственного или установленного образца;
* лица, получающие среднее профессиональное и (или) высшее образование (студенты, магистранты, аспиранты).
  1. Квалификация:
* педагогические работники, руководители (заместители руководителей, руководители структурных подразделений) профессиональных образовательных организаций, организаций дополнительного профессионального образования
* Исследователи в области социальных наук
  1. Наличие опыта профессиональной деятельности - да
  2. Предварительное освоение иных дисциплин/курсов /модулей не требуется

**4.Учебный план программы «Введение в Big Data: анализ естественного языка и взаимодействия пользователей»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль** | **Всего, час** | **Виды учебных занятий** | | |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | Модуль 1. Проектирование дизайна исследования в Big Data (основы методологии) | 12 | 2 | 2 | 8 |
| 2 | Модуль 2. Методы и технологии сбора, обработки и хранения данных в Big Data | 12 | 2 | 2 | 8 |
| 3 | Модуль 3. Social Network Analysis | 24 | 6 | 8 | 10 |
| 4 | Модуль 4. Методы и алгоритмы обработки естественного языка | 24 | 8 | 6 | 10 |
| **Итоговая аттестация** | |  | Выполнение контрольного задания | | |
| Всего | | **72** | 18/18/36 | | |

**5.Календарный план-график реализации образовательной** программы

(дата начала обучения – дата завершения обучения) в текущем календарном году, указания на периодичность набора групп (не менее 1 группы в месяц)

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Наименование учебных модулей** | **Трудоёмкость (час)** | **Сроки обучения** |
| 1 | Модуль 1. Проектирование дизайна исследования в Big Data (основы методологии) | 12 | 01.11 – 04.11 |
| 2 | Модуль 2. Методы и технологии сбора, обработки и хранения данных в Big Data | 12 | 05.11 – 08.11 |
| 3 | Модуль 3. Social Network Analysis | 24 | 09.11 – 14.11 |
| 4 | Модуль 4. Методы и алгоритмы обработки естественного языка | 24 | 14.11 – 21.11 |
| **Всего:** | | **72** | **01.11 – 21.11** |

**6.Учебно-тематический план программы «Введение в Big Data: анализ естественного языка и взаимодействия пользователей»**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | | | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | **Проектирование дизайна исследования в Big Data (основы методологии)** | **12** | 2 | 2 | 8 | Тест |
| 1.1 | **Введение в Big Data: перспективы, границы и возможности метода** | **3** | 1 | - | 2 | Тест |
| 1.2 | **Принципы организации аналитической работы с большими данными** | **3** | 1 | - | 2 | Тест |
| 1.3 | **Проектирование дизайна исследования с использованием технологий и методов Big Data** | **7** | - | 2 | 4 | Практическое задание |
| 2 | **Методы и технологии сбора, обработки и хранения данных в Big Data** | **12** | 2 | 2 | 8 | Практическое задание |
| 2.1 | **Источники открытых данных** | **4** | 1 | 1 | 2 | Тест |
| 2.2 | **Работа с API социальных сетей. Сбор открытой информации** | **8** | 1 | 1 | 6 | Практическое задание |
| 3 | **Social Network Analysis** | **24** | 5 | 8 | 10 | Практическое задание |
| 3.1 | **Social Network Analysis: сетевое взаимодействие между субъектами в социальных сетях** | **3** | 1 |  | 2 | Тест |
| 3.2 | **Структурный анализ взаимодействия в социальных сетях** | **10** | 2 | 4 | 4 | Практическое задание |
| 3.3 | **Анализ эгоцентрических сетей** | **5** | 1 | 2 | 2 | Практическое задание |
| 3.4 | **Оценка информационного распространения в социальных сетях.** | **5** | 1 | 2 | 2 | Практическое задание |
| 4 | **Методы и алгоритмы обработки естественного языка** | **24** | 6 | 8 | 10 | Практическое задание |
| 4.1 | **Текстовая аналитика с использованием платформы PolyAnalyst** | **3** | 2 |  | 1 | Тест |
| 4.2 | **Запуск и интерфейс платформы** | **3** | 1 | 1 | 1 | Практическое задание |
| 4.3 | **Основные узлы платформы** | **3** | 1 | 1 | 1 | Практическое задание |
| 4.4 | **Предобработка данных** | **4** | 1 | 1 | 2 | Практическое задание |
| 4.5 | **Использование языка XPDL** | **3** | 1 | 1 | 1 | Практическое задание |
| 4.6 | **Классификация текстов** | **4** | 1 | 1 | 2 | Практическое задание |
| 4.7 | **Кластеризация текстов** | **4** | 1 | 1 | 2 | Практическое задание |

**7. Учебная (рабочая) программа повышения квалификации «Введение в Big Data: анализ естественного языка и взаимодействия пользователей»**

**Модуль 1 Проектирование дизайна исследования в Big Data (основы методологии) (10 час.)**

**Тема 1.1 Введение в Big Data: перспективы, границы и возможности метода (2 час.)**

Демонстрируются различные основания определения концепта “Big Data”, показывается ретроспектива развития исследований, основанных на анализе больших данных.

Определение Big Data. Метод. Методология. Определение данных. Принципы работы с данными. Отличие Big Data от других направлений. Управление данными. Типовые кейсы в социальных науках. Методы решения. Карта компетенций. Литература. Интернет-ресурсы и онлайн-курсы.

**Тема 1.2 Принципы организации аналитической работы с большими данными (2 час.)**

Описываются ключевые принципы и специфика современной цифровой аналитики, структура аналитической деятельности. Дается представление о ключевых инструментах цифровой аналитики. Также описывается процесс организации аналитических групп и выделяются конкретные уровни и позиции аналитической работы.

**Тема 1.3 Проектирование дизайна исследования с использованием технологий и методов Big Data (6 час.)**

Описываются основные этапы проведения исследования с использованием технологий и методов сбора и обработки больших массивов данных. В ходе лекции будут представлены основные источники открытых данных, обзор инструментов анализа данных и примеры прикладных проектов на основе обработки открытой информации из социальных медиа.

**Модуль 2. Методы и технологии сбора, обработки и хранения данных в Big Data (12 час.)**

**Тема 2.1. Источники открытых данных (4 час.)**

Источники открытых данных. Типы больших пользовательских данных. Структурированные и неструктурированные данные.

**Тема 2.2. Работа с API социальных сетей. Сбор открытой информации (8 час.)**

В рамках практического занятия участники познакомятся с различными способами сбора открытых данных: методами работы с API (программный интерфейс приложения), парсингом открытых данных сайтов и порталов.

**Модуль 3. Введение в Social Network Analysis (23 час.)**

**Тема 3.1. Social Network Analysis: сетевое взаимодействие между субъектами в социальных сетях (3 час)**

Будет представлена методология анализа взаимодействия субъектов (людей, организаций) между собой в контексте различных социальных явлений. Будут проведены различия между структурным анализом, изучающим порядок взаимодействия между субъектами, позиционным анализом, показывающим наиболее влиятельных лиц в сети, а также динамическим анализом, который позволяет проанализировать, как изменяется взаимодействие между ними.

**Тема 3.2. Структурный анализ взаимодействия в социальных сетях (8 час.)**

Обучающиеся смогут построить сеть взаимодействия субъектов социальной сети и проанализировать её структуру, выделить ключевых лидеров при помощи программного продукта Gephi.

**Тема 3.3. Анализ эгоцентрических сетей (5 час)**

Различие между всей сеть и эго-сетью. Определение эго-сетей. Инструмент для извлечения эго-сетей. Визуализация эго-сетей.

**Тема 3.4. Оценка информационного распространения в социальных сетях (5 час)**

Слушатели узнают про диффузию и типы моделей распространения информации. Инфекционная и диффузная модели.

**Модуль 4. Методы и алгоритмы обработки естественного языка (24 час.)**

**Тема 4.1. Текстовая аналитика с использованием платформы PolyAnalyst**

Описание платформы для анализа текстов PolyAnalyst. Будут рассмотрены возможности платформы, преимущества ее использования, а также продемонстрированы примеры аналитических решений, разработанных с использованием платформы.

**Тема 4.2. Запуск и интерфейс платформы**

Запуск платформы и создание нового проекта. Слушатели познакомятся с интерфейсом, а также основными инструментами, с использованием которых строится аналитическое решение. Изучат операции по работе с узлами

**Тема 4.3. Основные узлы платформы**

Узлы в PolyAnalyst представляют собой определенную операцию либо набор операций. В данной части будут рассмотрены разделы, на которые подразделяются узлы. Особенности и спектр возможностей каждого из разделов а также их основное назначение.

**Тема 4.4 Предобработка данных**

Знакомство с основными этапами предобработки данных, необходимостью ее выполнения, а также влияние предобработки на результат. Сравнение выполнения некоторых этапов предобработки в языке программирования Python 3 и в платформе PolyAnalyst.

**Тема 4.5. Использование языка XPDL**

XPDL - («eXtended Pattern Definition Language», т.е. расширенный язык для определения текстовых паттернов) - это язык написания правил платформы PolyAnalyst, используемый для извлечения значимой информации из неструктурированных текстов. XPDL позволяет иерархически организовать PDL-запросы и находить сложные текстовые паттерны. Пользователи познакомятся с синтаксисом языка и изучат некоторые функции, необходимые для извлечения информации.

**Тема 4.6.Классификация текстов**

Классификация текстов — задача компьютерной лингвистики, заключающаяся в отнесении документа к одной из нескольких категорий на основании содержания документа. Рассмотрим реализацию алгоритмов классификации в платформе а также сравним реализацию некоторых алгоритмов на языке программирования Python 3.

**Тема 4.7. Кластеризация текстов**

Целью кластеризации является автоматическое выявление групп, семантически похожих документов среди заданного фиксированного множества. Существует огромное количество алгоритмов кластеризации. Основная идея большинства из них – объединить одинаковые последовательности в один класс или кластер на основе сходства. В текстовых данных, сравниваемыми составляющими служат последовательности слов и их атрибутов (например, вес слова в тексте, тип именованной сущности, тональность и пр.). Рассмотрим реализацию алгоритмов кластеризация в платформе PolyAnalyst.

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| **1.1** | Модуль 1 | Проектирование дизайна исследования с использованием технологий и методов Big Data | Построение дизайна исследования. Проблемная ситуация, цель, задача, источники данных, ожидаемые результаты |
| **1.2** | Модуль 2 | Сбор данных социальных сетей | Получение токена для использования API. Получение списка подписчиков сообщества социальной сети. Построение запроса в виде URL |
| 1.3 | Модуль 3 | Практика применения Social Network Analysis | Коррекция дизайна исследования с учетом применения SNA |
|  | Модуль 3 (тема 2) | Структурный анализ взаимодействия в социальных сетях | Выгрузка графа взаимодействия подписчиков сообщества в социальной сети. Визуализация графа. Выделение сетевых лидеров. Расчет сетевых статистик. Определение типа структуры. |
|  | Модуль 3 (тема 3) | Анализ эгоцентрических сетей | Выгрузка эго-графа пользователя социальной сети. Визуализация графа. Расчет статистических метрик. Интерпретация показателей. |
|  | Модуль 3 (тема 4) | Оценка информационного распространения в социальных сетях. | Определение модели распространения  Расчет параметров распространения. |
|  | Модуль 4 (Тема 1) | Основные узлы платформы и работа с ними | Рассмотрение выполнения основных операций с узлами на примере загрузки данных: добавление узла; настройка узла;  результат узла;  соединение узлов;  группировка/разгруппировка и удаление. |
|  | Модуль 4 (Тема 2) | Предобработка данных | Подготовка данных к дальнейшему анализу. Удаление дубликатов, индексирование, исправление орфографических ошибок, фильтрация и модификация данных. |
|  | Модуль 4 (Тема 3) | Поисковые запросы с использованием XPDL | Извлечение информации из неструктурированных текстов с помощью PDL запросов, поиск сложных текстовых паттернов |
|  | Модуль 4 (Тема 4) | Классификация | Выполнение классификации текстов с использованием обучающей выборки |
|  | Модуль 4 (Тема 5) | Кластеризация | Кластеризация текстовых данных, путем выявления семантической близости слов, словосочетаний и паттернов. |

**8.Оценочные материалы по образовательной программе**

**8.1. Вопросы тестирования по модулям**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ модуля** | **Вопросы входного тестирования** | **Вопросы промежуточного тестирования** | **Вопросы итогового тестирования** |
| **1.1** | Вопросы размещены на странице программы  <https://moodle.ido.tsu.ru/course/view.php?id=1197> | Вопросы размещены на странице программы  <https://moodle.ido.tsu.ru/course/view.php?id=1197> | Вопросы размещены на странице программы  <https://moodle.ido.tsu.ru/course/view.php?id=1197> |

**8.2.**  **Описание показателей и критериев оценивания, шкалы оценивания.**

Для оценки составляющих компетенции при **текущей аттестации** используется балльно-рейтинговая система шкалы оценок. Для определения фактических оценок каждого показателя по результатам сдачи практических работ обучающимися выставляются следующие баллы:

- результат, содержащий полный правильный ответ, полностью соответствующий требованиям критерия – 85 – 100 %;

- результат, содержащий неполный правильный ответ (степень полноты ответа – более 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия, – 75 – 84% от максимального количества баллов;

- результат, содержащий неполный правильный ответ (степень полноты ответа – до 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия – 60 - 74 % от максимального количества баллов;

- результат, содержащий неполный правильный ответ, содержащий значительные неточности, ошибки (степень полноты ответа – менее 60%) – до 60 % от максимального количества баллов;

- неправильный ответ (ответ не по существу задания) или отсутствие ответа, т.е. ответ, не соответствующий полностью требованиям критерия, – 0 % от максимального количества баллов;

**8.3.**  **Примеры контрольных заданий по модулям или всей образовательной программе**

* Опишите алгоритм предобработки текстовых данных
* Напишите отличия алгоритмов классификации и кластеризации текстовых данных
* Напишите различия между структурным и динамическим подходами в сетевом анализе
* Какие основные сетевые характеристики используются при анализе сетевого взаимодействия между людьми?
* Используя тестовый набор данных, рассчитать диаметр графа в пакете Gephi
* Используя тестовый набор данных, рассчитать плотность графа в пакете Gephi.
* Выгрузить данные о друзьях пользователя в социальной сети в формате graphml. Построить эго граф пользователя в пакете Gephi.
* Дано тестовое онлайн сообщество в социальной сети. Выгрузить данные о друзьях всех пользователей данного сообщества. Построить граф связей пользователей. Рассчитать среднюю длину связи, диаметр и плотность полученного графа
* Сформируйте отчет с рассчитанными параметрами и экспортируйте его в файл.

**8.4. Тесты и обучающие задачи (кейсы), иные практикоориентированные формы заданий.**

Задача – сформировать исследовательский проект на примере одной из тем:

* Анализ контента и дискурса онлайн-благотворительности
* Исследование цифрового качества жизни регионов
* Построение модели субъективного благополучия человека на основе цифровых следов
* Структура контента образовательных сообществ в социальных сетях
* Дискурсивный анализ девиантных сообществ в социальных сетях
* Структура взаимодействия участников девиантных сообществ

Структура представления исследовательского проекта:

* Проблемная ситуация, проблема
* Исследуемая гипотеза
* Источники данных
* Инструменты анализа данных
* Алгоритм анализа
* Ожидаемые результаты

**8.5.**  **Описание процедуры оценивания результатов обучения**

Формой оценки результатов освоения образовательного модуля выступает представление индивидуальных проектов.

Индивидуальный проект должен отражать:

* основания проведения исследовательского проекта (например, актуальность, новизна);
* гипотеза исследовательского проекта;
* необходимые методы исследовательского проекта;
* наличие сформированного в ходе проектной работы набора данных для реализации проекта;
* использование инструментов сбора данных, изученных в ходе образовательного модуля;
* использование методов математической статистики и их корректная интерпретация;
* использование методов машинного обучения и автоматического анализа текстов, социально-сетевого анализа для обработки набора данных;
* визуализация результатов обработки данных;
* релевантность полученных результатов для проверки гипотезы исследовательского проекта;
* примеры прикладного использования результатов исследовательского проекта;

К защите проекта допускаются слушатели, полностью выполнившие учебный план, не имеющие академической задолженности и представившие все необходимые для защиты документы.

Критерии финального оценивания:

Оценка «удовлетворительно» может быть поставлена при выполнении 60 процентов самостоятельной работы в курсе, выполнении проектной работы, отражающей недостаточное знание принципов работы программы и интерпретации методов и технологий сбора и обработки данных социальных сетей. Презентация работы выполнена с нарушением критериев, ответы на вопросы даны некорректно или не по существу.

Оценка «хорошо» может быть поставлена при выполнении 80 процентов самостоятельной работы в семестре, систематическом посещении занятий, выполнении проектной работы, отражающей недостаточное знание принципов работы программы и интерпретации методов и технологий сбора и обработки данных социальных сетей. Презентация работы выполнена согласно критериям, возможны недочеты в изложении содержания работы, затруднения при ответах на вопросы.

Оценка «отлично» может быть поставлена при выполнении 95-100 процентов самостоятельной работы в семестре, систематическом посещении занятий, выполнении проектной работы, отражающей достаточное знание принципов работы программы и интерпретации методов и технологий сбора и обработки данных социальных сетей. Презентация полностью отвечает всем критериям, ответы на вопросы преподавателя демонстрируют знание специфики релевантной области современного анализа социальных сетей.

**9.Организационно-педагогические условия реализации программы**

**9.1. Кадровое обеспечение программы**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Фамилия, имя, отчество (при наличии)** | **Место основной работы и должность, ученая степень и ученое звание (при наличии)** | **Ссылки на веб-страницы с портфолио (при наличии)** | **Фото в формате jpeg** | **Отметка о полученном согласии на обработку персональных данных** |
| **1** | Гойко Вячеслав Леонидович | Центр прикладного анализа данных НУ ТГУ, директор | https://persona.tsu.ru/Home/UserProfile/7256 | **D:\Рабочие документы\Настя\РЦДО\Персональные цифровые сертификаты\Заявка чисто\Делала во вторник\Программы\Программы\Гойко.png** | Получено |
| **2** | Кашпур Виталий Викторович | Кафедра социологии ТГУ, заведующий, канд. социол. наук | https://persona.tsu.ru/Home/UserProfile/1037 | **D:\Рабочие документы\Настя\РЦДО\Персональные цифровые сертификаты\Заявка чисто\Делала во вторник\Программы\Программы\Кашпур.png** | Получено |
| 4 | Мягков Михаил Георгиевич | Лаборатория экспериментальных методов в когнитивных науках, заведующий, PhD | https://persona.tsu.ru/Home/UserProfile/15040 | **D:\Рабочие документы\Настя\РЦДО\Персональные цифровые сертификаты\Заявка чисто\Делала во вторник\Программы\Программы\Мягков.png** | Получено |
| 5 | Мундриевская Юлия Олеговна | Центр прикладного анализа больших данных НУ ТГУ, мнс | https://persona.tsu.ru/Home/UserProfile/15977 | D:\Рабочие документы\Настя\РЦДО\Персональные цифровые сертификаты\Заявка чисто\Делала во вторник\Программы\Программы\Мундриевская.png | Получено |
|  | Губанов Александр Валерьевич | Центр прикладного анализа больших данных НУ ТГУ, аналитик | https://persona.tsu.ru/Home/UserProfile/27084 | D:\Рабочие документы\Настя\РЦДО\Персональные цифровые сертификаты\Заявка чисто\Делала во вторник\Программы\Программы\Губанов.png | Получено |
|  | Петров Евгений Юрьевич | Суперкомпьютерный центр НУ ТГУ, техник 2-й категории | https://persona.tsu.ru/Home/UserProfile/26676 | D:\Рабочие документы\Настя\РЦДО\Персональные цифровые сертификаты\Заявка чисто\Делала во вторник\Программы\Программы\Петров.png | Получено |

**9.2.Учебно-методическое обеспечение и информационное сопровождение**

|  |  |
| --- | --- |
| **Учебно-методические материалы** | |
| Методы, формы и технологии | Методические разработки, материалы курса, учебная литература |
| - | Электронный учебник Hanneman R. A. and Riddle M. (2005). Social network data, in: Hanneman R. A. and Riddle M. (2005) Introduction to social network methods. Riverside, CA: University of California, Riverside. URL: http://faculty.ucr.edu/~hanneman/nettext/ |
| - | Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press. (обновленная версия от 2016 года). Ch. 13-14 |

|  |  |
| --- | --- |
| **Информационное сопровождение** | |
| Электронные  образовательные ресурсы | Электронные  информационные ресурсы |
| - | https://habr.com/ru/hub/bigdata/ |

**9.3.Материально-технические условия реализации программы**

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования,  программного обеспечения |
| Практическое | редактор электронных таблиц MS Excel или OpenOffice Calc; |
| Практическое | текстовый редактор Notepad++ или любой другой, поддерживающий подсветку синтаксиса, переключение между разными кодировками и поиск с использованием регулярных выражений; |
| Практическое | среда разработки Jupyter Notebook (https://jupyter.org); |
| Практическое | язык программирования Python (https://www.python.org) ; |
| Практическое | веб браузер (Google Chrome, Opera, Mozilla) ; |
| Практическое | программный пакет Gephi 0.9.2. (https://gephi.org); |

**III.Паспорт компетенций (Приложение 2)**

ПАСПОРТ КОМПЕТЕНЦИИ

**Введение в Big Data: анализ естественного языка и взаимодействия пользователей**

Томский государственный университет

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Большие данные | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | профессиональная | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Знает:  - основные компетенции современных вычислительных систем  - основные ИКТ и программное обеспечения для решения задач интеллектуального анализа данных  - структуру и принципы реализации технологий и алгоритмов сбора, хранения и анализа данных  - особенности жизненного цикла исследовательского и прикладного проекта с использованием алгоритмов интеллектуального анализа данных  Умеет:  - проектировать научные или прикладные исследования с применением алгоритмов интеллектуального анализа данных  - использовать методы высокопроизводительных вычислительных технологий, современного программного обеспечения  - самостоятельно определять алгоритмы, необходимые для решения исследовательских и прикладных задач интеллектуального анализа данных  - использовать существующие программные решения для сбора, обработки и анализа данных  Владеет:  - инструментальными средствами высокопроизводительных вычислений в научной и практической деятельности  - моделями и средствами интеллектуального анализа данных  - средствами обработки естественного языка  - средствами сетевого анализа объектов и систем | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции  обучающегося | Индикаторы |
|  | | Начальный (Компетенция недостаточно  развита. Частично проявляет  навыки, входящие в состав  компетенции. Пытается,  стремится проявлять нужные  навыки, понимает их  необходимость, но у него не  всегда получается) | Знает  - основные компетенции современных вычислительных систем  - основные ИКТ и программное обеспечение для решения задач интеллектуального анализа данных социальных сетей  Умеет  использовать программные средства для решения однотипных задач анализа данных  Владеет  Программным обеспечением при решении прикладных задач интеллектуального анализа данных |
|  | | Базовый  (Уверенно  владеет навыками, способен,  проявлять соответствующие  навыки в ситуациях с  элементами  неопределённости сложности) | Знает  - средства и программное обеспечение для решения задач интеллектуального анализа данных  - структуру и принципы реализации технологий и алгоритмов сбора, хранения и анализа данных  Умеет  - проектировать научные или прикладные исследования с применением алгоритмов интеллектуального анализа данных в одной сфере  Владеет  - моделями и средствами интеллектуального анализа данных |
|  |  | | Профессиональный (Владеет  сложными навыками,  способен активно влиять на  происходящее, проявлять  соответствующие навыки в  ситуациях повышенной  сложности) | Знает  - алгоритмы, принципы и инструменты для решения задач интеллектуального анализа данных  Умеет  - проектировать научные или прикладные исследования с применением алгоритмов интеллектуального анализа данных для разных отраслей знания  - самостоятельно определять алгоритмы, необходимые для решения исследовательских и прикладных задач интеллектуального анализа данных  - использовать существующие программные решения для сбора, обработки и анализа больших данных  Владеет  - навыками самостоятельного освоения новых алгоритмов интеллектуального анализа данных социальных сетей  - средствами обработки естественного языка  - средствами сетевого анализа объектов и систем |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | * компетенции цифровой грамотности * способность решать прикладные задачи с использованием ИКТ | |
| 6. | Средства и технологии оценки | | Практические задания, защита проекта | |

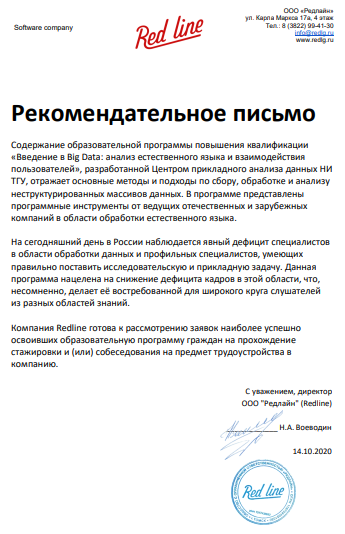
**VI.Иная информация о качестве и востребованности образовательной программы**

Нет

**V.Рекомендаций к программе от работодателей**: наличие двух писем от работодателей о рекомендации образовательной программы для реализации в рамках Государственной системы предоставления ПЦС на формирование у трудоспособного населения компетенций цифровой экономики с указанием востребованности результатов освоения программы в сфере деятельности соответствующих компаний и готовности к рассмотрению заявок наиболее успешно освоивших образовательную программу граждан на прохождение стажировки и (или) собеседования на предмет трудоустройства путем проставления отметки в профиле программы

Получены рекомендательные письма от двух компаний:

1. ООО «Редлайн»
2. ООО «Мегапьютер Интеллидженс»

****

****

**VI.Указание на возможные сценарии профессиональной траектории граждан** по итогам освоения образовательной программы (в соответствии с приложением)

* Переход в новую сферу занятости (освоение новой сферы / освоение смежных профессиональных областей)
* Развитие компетенций в текущей сфере занятости

**VII.Дополнительная информация**

-в основу проектирования программы положен компетентностный подход;

-использование информационных и коммуникационных технологий, в том числе современных систем технологической поддержки процесса обучения, обеспечивающих комфортные условия для обучающихся и преподавателей;

- применение современных образовательных технологий, инновационных методов обучения;

-решение реальных задач, стоящих перед обучающимися, соединение экспертного знания и опыта участников программы с их образовательной активностью;

-обучение в рамках образовательной программы реализуют преподаватели, прошедшие специальную подготовку по обозначенной тематике.

